动态加权网络中的演化社区发现算法研究 *

张高祯，张贤坤，苏　静，刘渊博

(天津科技大学 计算机科学与信息工程学院，天津 300457)

摘　要：在动态网络中发现社区结构是一个非常复杂而有意义的过程，可以更好地观察和分析网络的演化情况。针对动态加权网络中的社区发现问题，提出了一种结合历史网络社区结构的算法，叫做动态加权网络中的演化社区发现算法（ECDA）。该算法分为两步：结合历史社区和网络结构信息，计算当前时间跳的输入矩阵；然后通过该输入矩阵计算得到结合历史时间跳信息的社区划分结果。该算法有以下优点：可以自动发现动态加权网络中每个时间跳的社区结构；对网络结构的变化和社区结构的变化具有较高的敏锐性。在人工数据集和真实数据集中进行了实验，实验结果证明该算法可以有效地发现动态加权网络中的社区结构，与其他算法相比具有较好的竞争力。

关键词：动态网络；加权网络；社区发现；模块度

中图分类号：TP393

Research on evolutionary community discovery algorithm in dynamic weighted networks

Zhang Gaozhen, Zhang Xiankun, Su Jing, Liu Yuanbo

(*School of Computer Science & Information Engineering*, *Tianjin University of Science & Technology*, *Tianjin* 300457, *China*)

Abstract: In dynamic networks, detecting community structure is a very complex and meaningful process, which can better observe and analyze the evolution of the networks. For the community detection problem in dynamic weighted networks, this paper proposed an algorithm combining the community structure of the historical networks, called the evolutionary community discovery algorithm (ECDA) in dynamic weighted networks. The algorithm is divided into two steps: calculate the input matrix of the current timestep by combining the information of historical communities and network structure; and then calculate the result of community detection combining the historical timestep information through the input matrix. The algorithm has the following advantages: it can automatically discover the community structure of each timestep in the dynamic weighted network; the algorithm has a high sensitivity to the changes of network structure and the changes of community structure. And the experimental results show that the proposed algorithm can effectively detect the community structure in dynamic weighted networks, and it is quite competitive with other algorithms.

Key Words: dynamic networks; weighted networks; community detection; modularity

0　引言

由于复杂系统可以用复杂网络进行表示，例如：社会网络、邮件网络、生物网络以及技术网络等；因此得到了研究人员的广泛研究和探索。近年来，随着复杂网络研究的深入，人们发现很多实际网络都是由社区构成的[1]。复杂网络内部连接紧密的节点组成的集合称为社区。同一个社区内部的节点之间联系较为密切，而社区之间的节点联系则较为松散[2-7]。传统的社会网络分析方法专注于静态网络的研究，主要研究如何在网络中确定有意义的社区结构。但是在许多复杂网络中，实体间的交互是随时间动态变化的，而网络的变化必然导致其中的社区结构也随着时间有所变化[8]。网络存在的动态性增加了网络社区发现问题研究的复杂性，而针对静态网络的分析因为其不考虑动态网络中时间的参与而丢失了捕捉社区进化模式的机会，在一定程度上无法反映出动态网络中社区变化的动态性特点，影响网络分析结果的准确性。因此，关于动态网络社区发现的研究引发学者广泛关注，它已成为社区发现研究工作中一个热点问题。

动态网络演化社区发现的主要目标是发现不同时刻下的社区结构，主要研究的是发现动态社区结构的方法，并以此来揭示网络中隐含的不断动态变化的社区结构[9]。网络的动态性是指网络的成员与成员之间的相互关系会随时间的变化而变化。

基金项目：国家自然科学基金资助项目（61702367）

作者简介：张高祯（1993-），男，河南周口人，硕士研究生，主要研究方向为智能信息处理；张贤坤（1970-），男，安徽芜湖人，教授，博士，主要研究方向为智能信息处理；苏静（1977-），女，天津人，副教授，博士，主要研究方向为社会网络分析；刘渊博（1992-），女，山西太原人，硕士研究生，主要研究方向为智能信息处理.

国内外获取动态网络社区结构的算法可大致分为两类。第一类是基于两步分析策略的社区结构演化方法。首先，对社区网络数据进行时间跳划分，按照一定时间窗口集成数据，然后计算各个时间跳的社区结构，追踪相同社区在连续时间跳中的演化特征并且随时间的推移解释这些社区结构之间的差异。近年来，已经提出了许多检测演化社区结构的两步策略算法。Gergely 等人[10]开发了一种基于 clique percolation 的新算法，该算法首次允许在大范围内对重叠社区的时间依赖性进行研究，并以此为基础，描述了社区演化的基本关系。Hopcroft 等人[11]提出了一种基于不同时间跳快照的节点余弦相似度的跟踪聚类方法，以跟踪大型联网中不断变化的社区。Duan 等人[12]提出了一种社区挖掘算法，包括社区发现和动态加权有向图上的变化点检测。Takaffoli 等人[13]提出了一种社区匹配算法，以便随着时间的推移有效地识别和跟踪类似的社区。在这些研究中，社区发现和社区演化分别在两个阶段进行分析。历史的社区结构，包括与当前社区结构有关的有价值的信息，没有被考虑在内。

第二类方法考虑了历史网络社区结构信息的影响。根据动态网络缓慢演变的情况，Chakrabarti 等人[14]首先提出了演化聚类的框架。在这个框架下，演化聚类应该同时优化两个潜在冲突的条件：一个是任何时间跳的聚类应该尽可能保持对当前数据的忠实（或者一致性）；另一个是聚类结果不应该从一个时间跳到下一个时间跳发生显著的变化。Chi[15]提出了一个谱聚类进化算法，该算法通过优化由快照代价和时间代价组成的代价函数挖掘动态社区结构。快照代价衡量当前时刻下的聚类质量，时间代价衡量当前时刻和上一时刻之间聚类结果的相似程度。至此，动态社区发现问题可转换为优化代价函数的优化问题。Lin[16]提出了一个经典的 FacetNet 框架模型来实现动态社区发现，该算法采用了非负矩阵分解的方法同时分析社区和它们的演化，同时，FacetNet 采用的迭代求解方式可以保证算法最终收敛到一个局部最优解，但 FacetNet 存在的问题在于其划分结果的质量依赖于网络中社区的数目。Elham Havvaei 等人[17]提出一种动态复杂网络中发现重叠社区的博弈论方法。Guo 等人[18]提出了一种动态加权网络中的演化社区结构发现算法，但是该算法在初始社区、扩展社区以及合并社区的每个演化步骤中都需要人为的设定阈值，不同的阈值设定对算法的结果具有较大的影响。

在本文中，提出了一种动态加权网络中的演化社区发现算法（ECDA）。借鉴在大型网络中社区快速演变算法[19]，本文有效解决了在演化社区结构发现算法中对初始社区、扩展社区以及合并社区步骤阈值设定的人为干预问题。在本文提出的 ECDA 算法中，其可以自动发现动态加权网络中每个时间跳的社区结构。并且该算法对网络结构的变化和社区结构的变化具有较高的敏锐性。该算法在人工数据集和真实数据集中进行了实验，实验结果证明该算法可以有效地发现动态加权网络中的社区结构，与其他算法对比具有较好的竞争力。

1 相关概念

1.1 动态加权网络

一个具有 T 个时间跳的动态网络可以用一个网络序列 $G=\{G_1,G_2,...,G_T\}$ 表示。令 $G_t=(V_t,E_t)$ 表示在时间跳 t（$1\le t\le T$）时刻的网络结构，其中，$V_t=\{v_1^t,v_2^t,...,v_{N^t}^t\}$ 表示 G_t 网络中节点的集合，N^t 表示 G_t 网络中的节点数；$E_t=\{e_1^t,e_2^t,...,e_{M^t}^t\}$ 表示 G_t 网络中的边集，M^t 表示 G_t 网络中的边数；如果边集 E_t 不全为 1，那么 G_t 就是一个加权网络。如果在动态网络 G 中存在一个时间跳的网络 G_t 为加权网络，那么，G 就是一个动态加权网络。

一般地，在每一个时间跳 t 时刻的网络 G_t 都可以用一个邻接矩阵 $A^t=(w_{ij}^t)_{N^t\times N^t}$ 表示，其中，w_{ij}^t 表示节点 v_i^t 与节点 v_j^t 之间连接边的权重；如果节点 v_i^t 与节点 v_j^t 之间没有边相连，那么，$w_{ij}^t=0$。对于一个在时间跳 t，具有 N^t 个节点的网络 G_t，第 i 个节点的度定义为所有与节点 i 相连的边的权重之和，公式化表示为：

$$d_i^t=\sum_{j\in I^t} w_{ij}^t, \tag{1}$$

其中: $I^t=\{1,2,...,N^t\}$。

1.2 输入邻接矩阵

为了发现动态加权网络中的演化社区，必须要考虑以下的两个潜在条件：任何时间跳的聚类结果应尽可能保持对当前数据的忠实性（或者一致性）；聚类结果不应该从一个时间跳到下一个时间跳发生显著的变化。对于动态加权网络中，时间跳 t 时刻的加权网络 G_t（用邻接矩阵 $A^t=(w_{ij}^t)_{N^t\times N^t}$ 表示），本文定义一个输入矩阵 $U^t=(u_{ij}^t)_{N^t\times N^t}$，矩阵中的每一个元素定义如下[18]：

$$u_{ij}^t=\begin{cases} w_{ij}^t & \text{if } t=1 \\ (1-\alpha)w_{ij}^t+\alpha u_{ij}^{t-1}\delta_{ij}^{t-1} & \text{if } t\ge 2 \end{cases}, \tag{2}$$

其中: α 是一个用户自定义的参数，其范围是 $\alpha\in[0,1]$；

$$\delta_{ij}^{t-1}=\begin{cases} 1 & \text{if } C_i^{t-1}=C_j^{t-1} \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

其中: $C_i^{t-1}=C_j^{t-1}$ 表示在时间跳 $t-1$ 时刻节点 i 和节点 j 属于同一个社区。

本文对 u_{ij}^t 做如下解释。对于动态加权网络 G，在时间跳 $t=1$ 时刻的网络是初始加权网络，此时不存在 $t-1$ 时间跳网络；因此，此时的输入矩阵 U^1 就等于加权网络的邻接矩阵 A^1。当 $t\ge 2$ 时，此时的输入矩阵 U^t 要综合考虑当前时间跳 t 时刻加权网络的邻接矩阵 A^t 的信息和历史时间跳的加权网络信息以及历史社区结构信息。参数 α 是一个用于平衡上述两个条件的用户自定义参数，α 越大越侧重于强调历史网络和历史社区结构信息。随着 t 的增加，历史时间跳的网络和社区结构信息的影响会变得越来越小。

1.3 模块度

模块度也称模块化度量值，是由 Newman 和 Girvan[4]提出的用于衡量网络社区结构强度的方法。其定义为社区内部的总

边数和网络中总边数的比例减去一个期望值，该期望值是将网络设定为随机网络时同样的社区分配所形成的社区内部的总边数和网络中总边数的比例。为了方便计算，Newman[20]基于邻接矩阵重写了模块度函数。下面本文给出动态加权网络在时间跳 t 时刻的模块度：

$$Q^t = \frac{1}{2m^t} \sum_{i,j \in I^t} \left(u_{ij}^t - \frac{d_i^t d_j^t}{2m^t} \right) \delta_{ij}^t \tag{4}$$

其中：$d_i^t = \sum_{j \in I^t} u_{ij}^t$ 表示输入矩阵节点 i 的度；$d_j^t = \sum_{i \in I^t} u_{ij}^t$ 表示输入矩阵节点 j 的度；$m^t = \frac{1}{2} \sum_{i,j \in I^t} u_{ij}^t$ 表示输入矩阵 U^t 的边的权重和。

模块度的值小于 1，也可能为负值。模块度的值越大，说明划分的社区结构越稳定。当模块度值达到最大，接近 1 时，表明此时社区结构达到最强。当整个网络被看做一个社区时，此时的模块度值就是 0。对于一个网络划分，当社区内边的个数小于在随机网络中期望的边个数时，模块度的值可能为 0 或者为负值；在这种情况下，根据 Newman 和 Girvan 的定义，模块度的值为 0。对于大多数网络而言，模块度 Q 的值在区间 $[0.3, 0.7]$。

模块度增益公式 ΔQ 是由 Vincent 等人[19]在大型网络中社区快速演变算法中提出的。下面本文给出动态加权网络中，在时间跳 t 时刻的模块度增益 ΔQ^t 为

$$\Delta Q^t = \left[\frac{\sum_{in}^t + d_{i,in}^t}{2m^t} - \left(\frac{\sum_{tot}^t + d_i^t}{2m^t} \right)^2 \right] - \left[\frac{\sum_{in}^t}{2m^t} - \left(\frac{\sum_{tot}^t}{2m^t} \right)^2 - \left(\frac{d_i^t}{2m^t} \right)^2 \right] \tag{5}$$

其中：$\sum_{in}^t$ 表示在时间跳 t 时刻社区 C^t 内所有边的权重之和；$\sum_{tot}^t$ 表示在时间跳 t 时刻其他节点与社区 C^t 内节点相连的边的权重之和；$d_{i,in}^t$ 表示在时间跳 t 时刻节点 i 与社区 C^t 内节点相连的边的权重之和。本文通过计算节点分配前后的模块度增益 ΔQ^t，并记录 ΔQ^t 最大的那个邻居节点，如果 $\max \Delta Q^t > 0$，则把该节点分配到 ΔQ^t 最大的那个邻居节点所在的社区，否则保持不变。

2　算法描述

基于在第 1 章定义的输入矩阵，本文提出了一种动态加权网络中的演化社区发现算法（ECDA），其主要由两部分组成：计算输入矩阵和基于输入矩阵进行演化社区发现。动态加权网络中的演化社区发现算法的详细描述见算法 1。

算法 1：动态加权网络中的演化社区发现算法（ECDA）

输入：动态网络 $G = \{G_1, G_2, ..., G_T\}$，对应邻接矩阵 $\{A^1, A^2, ..., A^T\}$

输出：社区序列 $\{C_1^t, C_2^t, ..., C_{l^t}^t\}$，其中 $1 \le t \le T$，l^t 表示 t 时间跳的社区个数

1：$t \leftarrow 1$。

2：while $t \le T$ do

3：计算输入矩阵 U^t。

4：将输入矩阵 U^t 中的每个节点看做一个独立的社区，其中的社区数与节点个数相同。

5：对每一个节点 i，依次尝试把节点 i 分配到其邻居节点所在的社区。计算分配前后的模块度增益 ΔQ^t，并记录 ΔQ^t 最大的邻居节点。如果 $\max \Delta Q^t > 0$，则把节点 i 分配到 ΔQ^t 最大的邻居节点所在的社区；否则保持不变。

6：重复步骤 5，直到所有节点所属的社区不再发生变化。

7：对社区进行压缩，将所有在同一个社区的节点压缩成一个新的节点，社区内节点之间的边的权重转换为新节点自身的权重。社区间边的权重转换为新节点间的权重。

8：重复步骤 4~7，直到整个网络的模块度不在发生变化为止。

9：$t \leftarrow t+1$。

10：end

在步骤 5 中，本文只是想通过模块都增益来判断一个节点 i 是否可以加入社区 C^t 中，所以在实际中不必求出精确的模块度值，仅需要知道当前操作模块度增益值是否大于 0 即可，因此，本文得到模块度相对增益公式：

$$\Delta^{'} Q^t = d_{i,in}^t - \frac{\sum_{tot}^t \times d_i}{m^t}, \tag{6}$$

这里，模块度相对增益的值可能大于 1，但其不是真实的模块度增益值；然而它的正负结果表示了当前操作是否增加了模块度。因此，使用该模块度相对增益公式可以大大降低算法的时间复杂度。

算法复杂度分析。在计算输入矩阵时的计算复杂度是 $O(1)$；在计算基于输入矩阵进行演化社区发现时的计算复杂度是 $O(e)$；因此，可得在每个时间跳的时间复杂度为 $O(e)$。所以，整个动态网络的计算复杂度即为 $O(e^T)$。

3　实验结果与分析

在这里，本文使用了两个人工网络和一个真实网络（科学家研究合作网络数据集），用于评价动态加权网络中的演化社区发现算法（ECDA）的性能。本文使用了三个划分质量评价函数去测量得到的社区结构[18]。第一个函数是快照质量函数，用于评价当前社区划分结果关于当前网络数据的一致性。快照质量评价结果越高，表明社区划分结果越忠实于当前时间跳网络数据。快照质量函数如下：

$$SQ^t = \frac{1}{2m^t} \sum_{i,j \in I^t, i \ne j} \left(w_{ij}^t - \frac{d_i^t d_j^t}{2m^t} \right) \delta_{ij}^t . \tag{7}$$

这里，d_i^t 表示 t 时间跳的网络 A^t 中节点的度；d_j^t 表示 t 时间跳的网络 A^t 中节点 j 的度；m^t i 表示 t 时间跳的网络 A^t 中所有边的权重和；其他参数含义同上。

第二个函数是历史质量函数，用于测量当前社区划分结构关于历史网络信息和社区结构信息拟合效果的瞬时平滑性。若得到的历史质量评价结果较高，说明此时的社区划分结构在上

一时间跳到当前时间跳之间没有发生显著的变化。把当前时间跳的社区划分结果应用到上一时间跳的数据中，得到历史质量评价函数：

$$HQ^t = \frac{1}{2m^{t-1}} \sum_{i,j \in I^{t-1}, i \neq j} \left(w_{ij}^{t-1} - \frac{d_i^{t-1} d_j^{t-1}}{2m^{t-1}} \right) \delta_{ij}^t . \tag{8}$$

这里，d_i^{t-1} 表示 $t-1$ 时间跳的网络 A^{t-1} 中节点 i 的度；d_j^{t-1} 表示 $t-1$ 时间跳的网络 A^{t-1} 中节点 j 的度；m^{t-1} 表示 $t-1$ 时间跳的网络 A^{t-1} 中所有边的权重和；其他参数含义同上。

最后一个函数是整体质量函数，用于评价平衡两个潜在冲突条件后的网络整体划分质量。整体质量评价结果越高，表明网络社区划分结果整体质量越高。整体质量函数如下：

$$TQ^t = \frac{1}{2m^t} \sum_{i,j \in I^t, i \neq j} \left(u_{ij}^t - \frac{d_i^t d_j^t}{2m^t} \right) \delta_{ij}^t . \tag{9}$$

这里，d_i^t 表示 t 时间跳的输入矩阵 U^t 中节点 i 的度；d_j^t 表示 t 时间跳的输入矩阵 U^t 中节点 j 的度；m^t 表示 t 时间跳的输入矩阵 U^t 中所有边的权重和；其他参数含义同上。

在这里，本文实现并改进了 Guo 等人[18]提出的动态加权网络中的演化社区结构发现算法（ECSD）。本文对 ECSD 算法中扩展社区、合并社区步骤进行了优化；本文在 ECSD 扩展社区时选择节点对社区模块度贡献最大且大于阈值的社区进行加入；在 ECSD 算法合并社区时选择两两合并时模块度最大且大于设定的阈值的两个社区进行合并。并与本文提出的算法（ECDA）进行了对比实验，从而证明了本文提出的动态加权网络中的演化社区发现算法（ECDA）的有效性和竞争力。

本文提出的算法（ECDA）还与 ECSD 算法、PDG 算法[21]和 InfoMap 算法[22]进行了每个时间跳社区划分质量的模块度对比实验。由于 PDG 算法和 InfoMap 只能用于非加权的网络，所以本文在这里只在非加权的人工网络和真实网络中进行了每个时间跳的模块度对比实验。通过分析其他文献的模块度评价对比指标，发现他们的模块度质量评价指标就是本文上述的三个评价函数中的快照评价（SQ）；为了保持与其他文献说法上的一致性，本文这里也采用模块度质量评价的叫法。

3.1　人工网络

在这里，本文使用了两个人工网络用于验证本文提出算法（ECDA）的有效性，它们分别是 Guo 等人使用的小规模动态加权网络数据集和使用 LFR 生成的人工网络数据集。

3.1.1 小规模动态加权网络

在这个小规模动态加权网络[18]中，每一个时间跳中的节点和边都是不断变化的，得到的社区结构也是随着时间不断变化的。本文使用该人工网络测试所提出算法（ECDA）的有效性。初始的动态加权网络如图 1 所示。

从图 1 中，可以观察到在时间跳 1 时的网络中有 7 个节点，它们可以被划分成 2 个社区。然而，在时间跳 2 时的网络节点数、边数和边权重都发生了变化；例如，节点 2 与节点 3 之间的边消失了，新增了一个节点 8 等。在时间跳 3 时的网络的边数以及边权重也发生了变化。在时间跳 4 时的网络可以明显被划分成 3 个社区。值得注意的是，在时间跳 2 时的网络中的节点 5 可以被划分到两个社区中（节点 1，2，3，4 组成的社区和节点 6，7，8 组成的社区）；但是参考时间跳 1 时的网络划分，应该把节点 5 划分到第二个社区（节点 6，7，8 所组成的社区）中；在这种情况下，就可以看到历史网络和社区信息就变得非常重要。

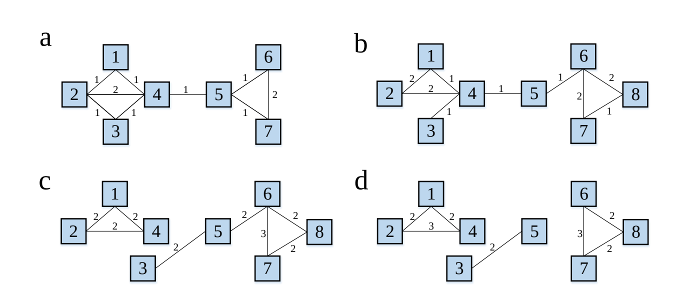


图 1　初始动态加权网络，

其中 a，b，c 和 d 分别表示在时间跳 1，2，3 和 4 时刻的网络结

表 1　在人工动态加权网络中每个时间跳的 SQ、HQ 和 TQ 的评价结果

质量评价	时间跳 1	时间跳 2	时间跳 3	时间跳 4
SQ	0.550	0.568	0.642	0.738
HQ	--	0.550	0.509	0.642
TQ	0.550	0.578	0.626	0.737

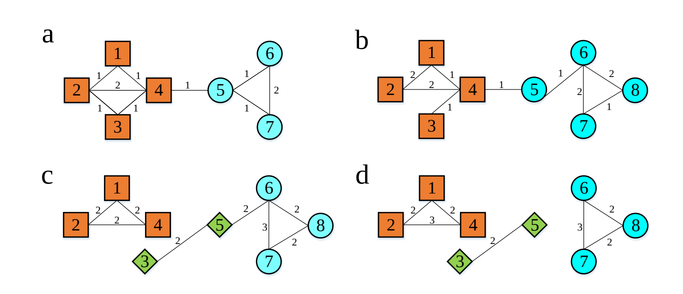


图 2　在人工动态加权网络中使用 ECDA 算法得到的划分结果，其中 a，b，c 和 d 分别代表在时间跳 1，2，3 和 4 时的社区划分结果

使用 ECDA 算法在这个人工动态加权网络中的划分结果如图 2 所示。本文在表 1 中给出了快照评价结果（SQ）、历史评价结果（HQ）和整体评价结果（TQ）的详细信息。在时间跳 1 时，由于没有历史网络信息和历史社区结构信息，所以没有历史评价结果；而且由第二部分的介绍可知，此时的快照质量评价就等于整体质量评价。同时，本文给出了 ECDA 算法在该人工网络中对比 ECSD 算法的实验结果，如图 3 所示，证明了本文提出的算法的有效性和竞争力。

3.1.2 LFR 人工网络

本文使用 LFR 生成了四个数据集，然后以这四个数据集为四个时间跳进行实验，生成网络的参数设置如表 2 所示。其中，

N 表示节点数；k 节点的平均度；maxk 表示节点的最大度；mu 表示混合参数；minc 表示最小社区尺寸；maxc 表示最大社区尺寸。自上而下分别是 LFR 生成四个数据集的参数设置。

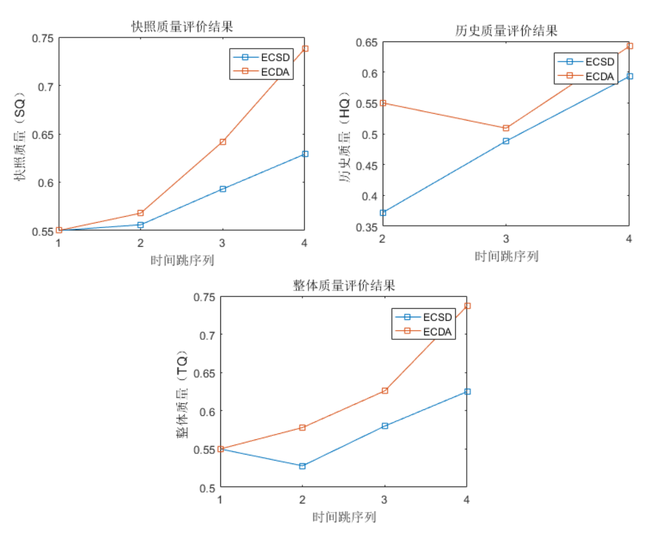


图 3 人工动态加权网络的每个时间跳网络中,本文提出的算法(ECDA)与 ECSD 算法在三个质量评价结果的对比

表 2 使用 LFR 生成四个网络的相关参数设置值

LFR	N	k	maxk	mu	minc	maxc
LFR1	1000	20	30	0.2	30	60
LFR2	1000	25	30	0.2	30	60
LFR3	1200	25	30	0.2	30	60
LFR4	1100	20	30	0.2	30	60

本文使用提出的动态加权网络中的演化社区发现算法（ECDA）和对比的 ECSD 算法分别在该数据集中进行了实验。在表 3 中给出了本文的算法（ECDA）在 LFR 人工网络中的快照评价结果（SQ）、历史评价结果（HQ）和整体评价结果（TQ）的详细信息。

表 3 在 LFR 人工网络中每个时间跳的 SQ、HQ 和 TQ 的评价结果

质量评价	LFR1	LFR2	LFR3	LFR4
SQ	0.688	0.746	0.741	0.744
HQ	--	-0.002	0.005	-0.001
TQ	0.688	0.641	0.653	0.629

从表中可以看到，历史质量评价结果（HQ）有负值的情况；原因是本文使用 LFR 生成的各个网络之间没有必然的联系，也就是网络之间的变化较大，得到的社区结构变化也比较大；因此，得到负值情况的历史质量评价结果（HQ）。

同时，本文给出了 ECDA 算法在该 LFR 网络中对比 ECSD 算法的实验结果，如图 4 所示，证明了本文提出的算法的性能。

从图 4 的质量评价结果中可以看出，本文提出的动态加权网络中的演化社区发现算法（ECDA）在快照质量（SQ）和整体质量（HQ）评价结果中具有较好的结果。因为历史质量（HQ）评价结果显示的是当前社区划分结构关于历史网络信息和社区结构信息拟合效果的瞬时平滑性；如果网络结构和社区结构在不同时间跳之间发生较大波动，那么得到的历史质量评价结果就会较低。因此，在 LFR 网络的历史评价（HQ）结果中，本文得到的结构低于 ECSD 算法，证明网络和社区结构在不同 LFR 生成网络之间发生了较大变化，说明了本文的算法（ECDA）对网络结构的变化和社区结构变化的敏锐性。

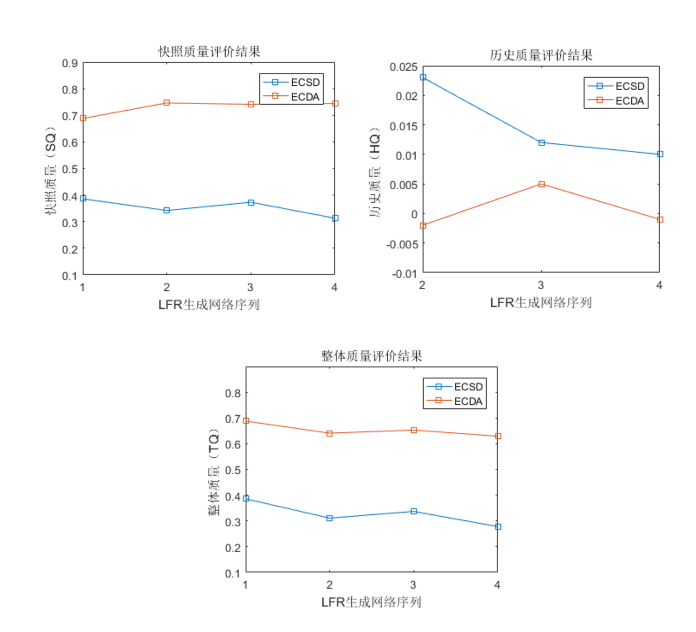


图 4 在 LFR 网络中,动态加权网络中的演化社区发现算法(ECDA)与 ECSD 算法在三个质量评价结果的对比

此外，在该 LFR 网络中，本文提出的 ECDA 算法还与 PDG 算法、InfoMap 算法和 ECSD 算法进行了每个时间跳社区划分结果的模块度质量对比实验。实验结果如图 5 所示。

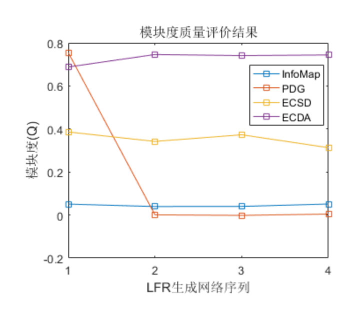


图 5 在 LFR 网络中，ECDA 算法与 PDG 算法、InfoMap 算法和 ECSD 算法的社区划分结果质量的模块度对比

从图 5 中，可以看到，本文提出的算法在该 LFR 网络中整体具有较好的划分结果。除了在第一个时间跳本文的算法的模块度略低于 PDG 算法之外；在其他 3 个时间跳中，本文提出的 ECDA 算法都得到了较好的模块度质量。说明本文提出的 ECDA 算法具有较好的社区划分性能。

3.2 真实网络

本文使用了一个真实网络数据集（科学家研究合作网络数据集[18]）用于说明本文提出算法（ECDA）的有效性和稳定性。

科学家研究合作网络是“985 工程”我国政府为建设若干所世界一流大学和一批国际知名的高水平研究型大学而实施的

高等教育建设工程大学之间的校际合作论文网络。“985 工程”名称源自 1998 年 5 月 4 日，江泽民在北京大学百年校庆上建设世界一流大学的讲话。最初入选 985 工程的高校有九所，被称九校联盟，截至 2011 年年末，985 工程共有 39 所高校。本文中收集了 1996 年至 2003 年中国“985 项目”大学之间在 WoS 上的校际合作论文。

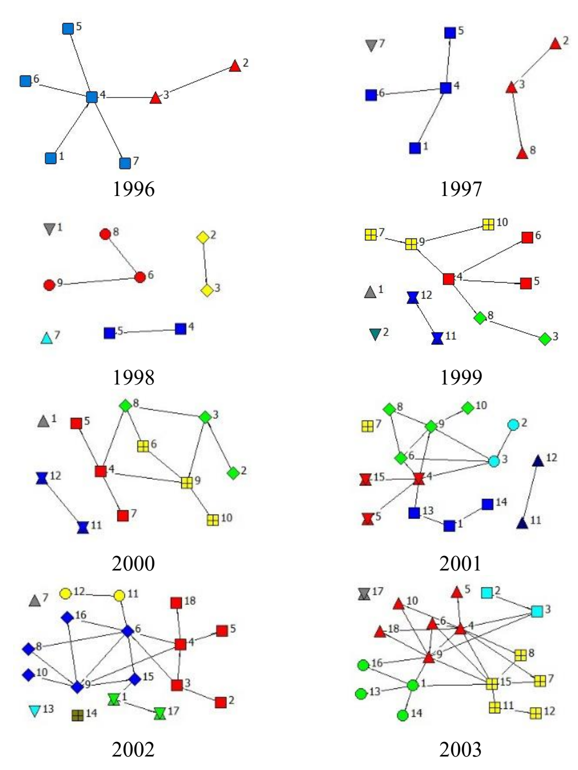


图 6　通过 ECDA 算法在 1996 年至 2003 年科学家研究合作网络中的社区划分结果

这里，本文按照每一年作为一个时间跳来组织该动态网络，使用本文上文提出的算法得到的社区划分结果如图 6 所示。

在表 4 中给出了本文的算法（ECDA）在科学家研究合作网络中的快照评价结果（SQ）、历史评价结果（HQ）和整体评价结果（TQ）的详细信息。在时间跳 1996 年时，由于没有历史网络信息和历史社区结构信息，所以没有历史评价结果；而且通过第二部分的介绍可知，此时的快照质量评价就等于整体质量评价。

表 4　在科学家研究合作网络中每个时间跳的 SQ、HQ 和 TQ 的评价结果

质量评价	1996	1997	1998	1999
SQ	0.444	0.660	0.781	0.602
HQ	--	0.389	0.280	0.063
TQ	0.444	0.627	0.678	0.550
质量评价	2000	2001	2002	2003
SQ	0.475	0.479	0.449	0.374
HQ	0.352	0.368	0.291	0.209
TQ	0.481	0.488	0.455	0.367

在这里，同样本文给出了 ECDA 算法在该科学家研究合作网络中对比 ECSD 算法的实验结果，如图 7 所示，说明了本文提出的算法的性能。

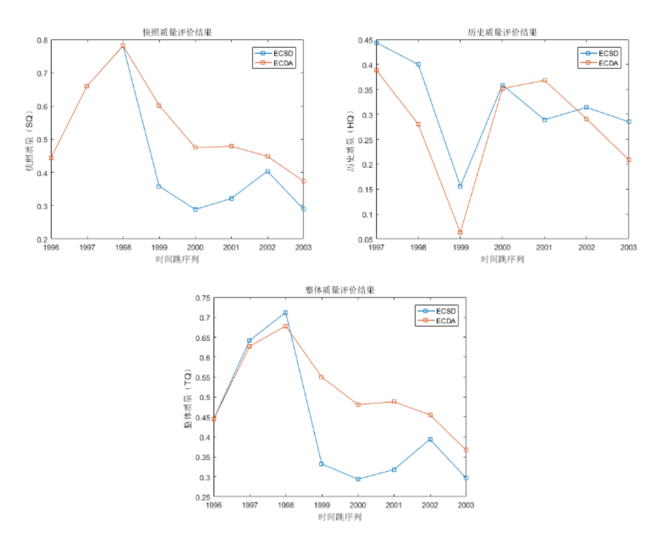


图 7 在科学家研究合作网络的每个时间跳网络中，本文提出的算法（ECDA）与 ECSD 算法在三个质量评价结果的对比

从上图的评价结果中可以看出，本文提出的 ECDA 算法在快照质量（SQ）和整体质量（HQ）评价结果中具有较好的结果。因为历史质量（HQ）评价结果显示的是当前社区划分结构关于历史网络信息和社区结构信息拟合效果的瞬时平滑性；如果网络结构和社区结构在不同时间跳发生较大波动，那么得到的历史质量评价结果就会较低。因此，在科学家研究合作网络的历史评价（HQ）结果中，本文得到的结果低于 ECSD 算法，说明网络和社区结构在不同时间跳之间发生了较大变化，验证了本文的算法（ECDA）对网络结构的变化和社区结构的变化更加敏锐。

这里，本文同样在该科学家研究合作网络数据集中用本文提出的 ECDA 算法与 PDG 算法、InfoMap 算法和 ECSD 算法进行了每个时间跳社区划分结果质量的模块度对比实验。实验结果如图 8 所示。

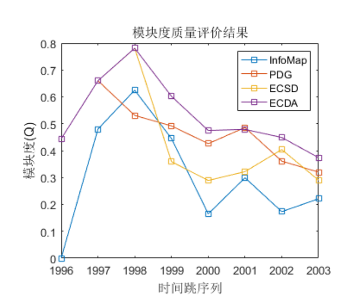


图 8　在科学家研究合作网络中,本文提出的 ECDA 算法与 PDG 算法、InfoMap 算法和 ECSD 算法的社区划分结果质量的模块度对比

从图 8 中可以看出，本文提出的 ECDA 算法具有较好的社区划分结果。除了在时间跳 2001 时得到的模块度略低于 PDG 算法外，其他都具有较好的划分结果。验证了本文提出的 ECDA 算法与其他算法相比具有较好的竞争力。

4 结束语

本文提出了一种动态加权网络中的演化社区发现算法（ECDA）。该算法可以自动发现动态加权网络中每个时间跳的社区结构。该算法在人工数据集和真实数据集中进行了实验，实验结果证明该算法可以有效地发现动态加权网络中的社区结构，与其他算法对比具有较好的竞争力。而且实验证明该算法对动态加权网络结构的变化和社区结构的变化更具有敏锐性。但是，该算法依然存在不足，该算法不能用在有向网络中；而且如何分析在不同时间跳社区结构具体的变化情况依然是一个研究热点问题。

参考文献：

[1] 毛蓝. 基于自相似特性的网络业务流的研究 [D]. 无锡: 江南大学, 2007.

[2] Freeman L C. The sociological concept of‘group’: an empirical test of two models [J]. American Journal of Sociology, 1992, 98 (1): 152.

[3] Porter M A, Onnela J P, Mucha P J. Communities in networks [J]. Notices American Mathematical Society, 2009, 56 (1082): 1.

[4] Girvan M, Newman M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences, 2002, 99 (12): 7821.

[5] Hopcroft J, Khan O, Kulis B, et al. Natural communities in large linked networks [C]// Proc of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2003: 541.

[6] Strogatz S H. Exploring complex networks [J]. Nature, 2001, 410 (6825): 268.

[7] Giatsoglou M, Vakali A. Capturing social data evolution using graph clustering [J]. IEEE Educational Activities Department, 2013, 17 (1): 74.

[8] 周旭. 复杂网络中社区发现算法研究 [D]. 吉林: 吉林大学, 2016.

[9] 王莉, 程学旗. 在线社会网络的动态社区发现及演化 [J]. 计算机学报, 2015, 38 (2): 219.

[10] Palla G, Barabsi A L, Vicsek T. Quantifying social group evolution [J]. Nature, 2007, 446 (7136): 664.

[11] Hopcroft J, Khan O, Kulis B, et al. Tracking evolving communities in large linked networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101 (Suppl 1): 5249.

[12] Duan D, Li Y, Jin Y, et al. Community mining on dynamic weighted directed graphs [C]// Proc of ACM International Workshop on Complex Networks Meet Information & Knowledge Management. 2009: 11.

[13] Takaffoli M, Sangi F, Fagnan J, et al. Community evolution mining in dynamic social networks [J]. Procedia-Social and Behavioral Sciences, 2011, 22 (22): 49.

[14] Chakrabarti D, Kumar R, Tomkins A. Evolutionary clustering [C]// Proc of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2006: 554.

[15] Chi Y, Song X D, Zhou D, et al. Evolutionary spectral clustering by incorporating temporal smoothness [C]// Proc of the 13th International Conference on Knowledge Discovery and Data Mining. 2007: 153.

[16] Lin Y R, Zhu S, Sundaram H, et al. Analyzing communities and their evolutions in dynamic social networks [J]. ACM Trans on Knowledge Discovery from Data, 2009, 3 (2): 1.

[17] Alvari H, Hajibagheri A, Sukthankar G. Community detection in dynamic social networks: a game-theoretic approach [C]// Proc of IEEE//ACM International Conference on Advances in Social Networks Analysis and Mining. 2014: 101.

[18] Guo C, Wang J, Zhang Z. Evolutionary community structure discovery in dynamic weighted networks [J]. Physica A Statistical Mechanics & Its Applications, 2014, 413 (11): 565.

[19] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics-Theory and Experiment, 2008, 2008 (10): 155.

[20] Newman M E J, Girvan M. Finding and evaluating community structure in networks [J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2004, 69 (2 Pt 2): 026113.

[21] 许宇光, 蒋飞, 朱恩强, 等. 基于个体稳定度博弈的动态社区发现算法研究 [J]. 电子与信息学报, 2017, 39 (4): 763.

[22] Rosvall M, Bergstrom C T. Maps of random walks on complex networks reveal community structure [J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 105 (4): 1118.